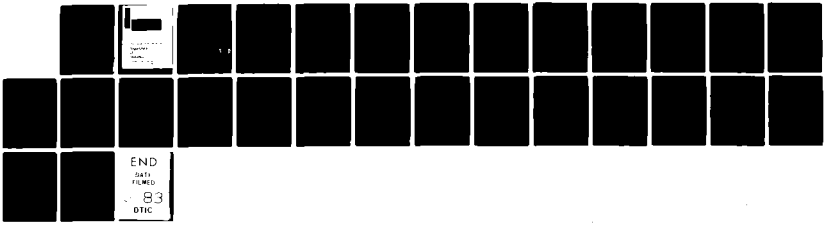
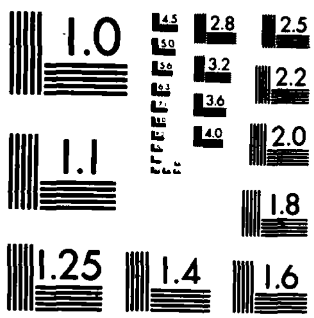
AD-A123 912 USE OF LOG-LINEAR MODELS IN CLASSIFICATION PROBLEMS(U) 1/
FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS
T C REDMAN DEC 81 FSU-STATISTICS-M592 N00014-76-C-0394

UNCLASSIFIED F/G 12/1 NL



END
DATE FILMED
83
DTIC



USE OF LOG-LINEAR MODELS IN CLASSIFICATION PROBLEMS

by

Thomas C. Redman

The Florida State University
Department
of
Statistics
Tallahassee, Florida
32306

DTIC A

USE OF LOG-LINEAR MODELS IN CLASSIFICATION PROBLEMS

by

Thomas C. Redman

FSU Statistics Report No. M592
ONR Technical Report No. 154

December, 1981

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Research supported in part by the Office of Naval Research under Contracts N00014-76-C-0394 and N00014-80-C-0093 and in part by the National Institute of Environmental Health Sciences under Grant 5 T32 ES07011. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Use of Log-Linear Models in Classification Problems[1]

by

Thomas C. Redman[2]
The Florida State University
Tallahassee, Florida 32306

In this paper we consider use of some special log-linear models and minimum δ estimation in the multivariate classification problem, posed by Martin and Bradley (1972). We first define these models, called log-difference models, and show that the minimum risk classification rule depends only on a certain subset of the new parameters. We then review minimum δ estimation, in particular the minimum δ estimator, the approximate minimum δ estimator, and their existence properties. Two examples are worked. The first involves detergent preference and illustrates how extensions to the case in which not all variables are dichotomous may be obtained through the use of orthogonal polynomials. The second example involves infant hypoxic trauma, and many cells are empty. The existence conditions are used to find a model for which estimates of cell frequencies exist and are in good agreement with the observed data.

Key Words: Classification problem, log-difference models, minimum δ estimation, existence.

---

[1]Research supported in part by the Office of Naval Research under Contracts N00014-76-C-0394 and N00014-80-C-0093 and in part by the National Institute of Environmental Health Sciences under Grant 5 T32 ES07011. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2]Now a Member of Technical Staff, Bell Laboratories, Holmdel, New Jersey, 07733.

1. Introduction

This paper serves the dual purposes of extending the classification problem considered by Martin and Bradley (1972), and of illustrating the uses of minimum δ estimation (Redman, 1981).

Martin and Bradley (hereafter denoted MB) consider the problem of classifying individuals from L populations, $\Pi^{(1)}, \ldots, \Pi^{(L)}$, into J categories, $C_1, \ldots, C_J$, on the basis of a random vector $\underset{\sim}{Z}$ which consists of I dichotomous variates. They reparameterize the 2^I possible state probabilities as

$$\pi^{(\ell)}(\underset{\sim}{z}) = \pi(\underset{\sim}{z})[1 + h(\underset{\sim}{a}^{(\ell)}, \underset{\sim}{z})], \tag{1}$$

where $\pi^{(\ell)}(\underset{\sim}{z})$ denotes the probability of state $\underset{\sim}{z}$ for the ℓ^{th} population, $\pi(\underset{\sim}{z})$ denotes the probability of state $\underset{\sim}{z}$ for a well-defined composite population, and $h(\underset{\sim}{a}^{(\ell)}, \underset{\sim}{z})$ is expressed in terms of 2^I orthogonal polynomials, the coefficients in $\underset{\sim}{a}^{(\ell)}$ being specific to the ℓ^{th} population. Models arise through the approximation of $h(\underset{\sim}{a}^{(\ell)}, \underset{\sim}{z})$ in (1) by a set of low-order polynomial terms, $h_s(\underset{\sim}{a}^{(\ell)}, \underset{\sim}{z})$. Thus, models for $\pi^{(\ell)}(\underset{\sim}{z})$ are of the form

$$\pi^{(\ell)}(\underset{\sim}{z}) = \pi(\underset{\sim}{z})[1 + h_s(\underset{\sim}{a}^{(\ell)}, \underset{\sim}{z})]. \tag{2}$$

In this paper we generalize the problem. We assume that the various levels or categories for the I variates define k states, which are labeled consecutively. Thus, while MB define cells in their tables by an I-vector $\underset{\sim}{Z}$, we simply take Z to be a variable which may take on values 1, ..., k.

In Section 2 we propose use of a model, called the log-difference model, in the classification problem. As with the difference model of

MB, the log-difference model features some parameters that are general to all L populations and some that are specific to individual populations.

In Section 3 we review the classification problem in detail. In particular the minimum risk classification rule is shown to depend on those parameters specific to individual populations only.

In Section 4 the minimum δ and approximate minimum δ estimation procedures are introduced. These were developed due to the lack of convenient conditions for the existence of maximum likelihood estimates in sparse data situations. Convenient conditions for the minimum δ estimator have been developed and are stated here. The new estimators have been shown to be asymptotically equivalent to the maximum likelihood estimator, so should yield good results when sample sizes are large. Full details may be found in Redman (1981).

Examples are given in Sections 5 and 6. The first is designed to illustrate the use of the log-difference model when one of the variates has ordered categories. The sample size is large, and the maximum likelihood, minimum δ and approximate minimum δ estimates are nearly equal. In the second example data are sparse, and the conditions stated in Section 4 are used to find an adequate model for which a minimum δ estimate exists.

2. The Log-Difference Model

Before proceeding with the development of the log-difference model, it is necessary to introduce some notation. The notation is similar to that used in Redman (1981), but is somewhat simpler due to the structure necessary in the classification problem. Throughout, we consider two sampling situations: the independent samples case, in which independent

samples of size $N^{(\ell)}$ are available from each population, and the single sample case, in which the data arise from a single sample of size N from all L populations. We shall use an index s = L for the independent samples case and s = 1 for the single sample case.

Let $\underset{\sim}{n}$ be the random (kL)-vector whose elements denote the number of observations which fall in each of the k cells of the L populations. Thus $\underset{\sim}{n}' = \{[\underset{\sim}{n}^{(1)}]', \ldots, [\underset{\sim}{n}^{(L)}]'\}$, where

$$[\underset{\sim}{n}^{(\ell)}]' = [n_1^{(\ell)}, \ldots, n_k^{(\ell)}], \quad \ell = 1, \ldots, L,$$

and $n_i^{(\ell)}$ denotes the number of observations for state i of the ℓ^{th} population. In the independent samples case, we assume the multinomial distribution,

$$\underset{\sim}{n}^{(\ell)} \sim \text{Mult}\left[N^{(\ell)}, \underset{\sim}{\pi}^{(\ell)}; \pi_i^{(\ell)} > 0, \sum_{i=1}^{k} \pi_i^{(\ell)} = 1\right], \quad \ell = 1, \ldots, L,$$

and define $\underset{\sim}{p}' = [[\underset{\sim}{p}^{(1)}]', \ldots, [\underset{\sim}{p}^{(L)}]']$ through $p_i^{(\ell)} = n_i^{(\ell)}/N^{(\ell)}$. In the single sample case, we assume

$$\underset{\sim}{n} \sim \text{Mult}\left[N, \underset{\sim}{\pi}, \underset{\sim}{\pi}' = \{[\underset{\sim}{\pi}^{(1)}]', \ldots, [\underset{\sim}{\pi}^{(L)}]'\}; \pi_i^{(\ell)} > 0, \sum_{\ell=1}^{L} \sum_{i=1}^{k} \pi_i^{(\ell)} = 1\right]$$

with $p_i^{(\ell)} = n_i^{(\ell)}/N$. Note that, in the independent samples case, s = L, $\pi_i^{(\ell)}$ is interpreted as the probability of state i for population ℓ, and, in the single sample case, s = 1, as the joint probability of state i and population ℓ.

In an analogous fashion, we define $\underset{\sim}{\gamma}$ and $\underset{\sim}{\gamma}^{(\ell)}$, $\ell = 1, \ldots, L$, through $\gamma_i^{(\ell)} = \log \pi_i^{(\ell)}$, i = 1, ..., k. A log-linear model is specified by m orthonormal constraints on $\underset{\sim}{\gamma}$, $\underset{\sim}{B}_1\underset{\sim}{\gamma} = \underset{\sim}{0}_m$, where $\underset{\sim}{B}_1\underset{\sim}{A}(s) = \underset{\sim}{0}_{m\times s}$ and $\underset{\sim}{A}(s) = \underset{\sim}{1}_{kL}$ when s = 1 and $\underset{\sim}{A}(s) = \begin{bmatrix} \underset{\sim}{1}_k & & \underset{\sim}{0} \\ & \ddots & \\ \underset{\sim}{0} & & \underset{\sim}{1}_k \end{bmatrix}_{kL\times L}$ when s = L.

We now describe a two-stage reparameterization of $\underset{\sim}{\gamma}$ which is useful in the classification problem. In the first stage, (L + 1) sets of k parameters are defined; one set is general to all populations, and each of the others is specific to a given population. The set specific to population L is redundant in that it is a linear function of the other sets and is not considered further. Each of the remaining L sets is further reparameterized in the second stage. The motivation behind this second reparameterization is in definition of the log-difference model in which certain linear functions of this final set may be assumed to be zero. This permits a reduction in the number of independent parameters to be estimated and is of particular importance when data are sparse.

Define k-vectors $\underset{\sim}{\gamma}_g$, general parameters, and $\underset{\sim}{\Delta}^{(\ell)}$, parameters specific to population ℓ, through

$$\underset{\sim}{\gamma}^{(\ell)} = \underset{\sim}{\gamma}_g + \underset{\sim}{\Delta}^{(\ell)}, \quad \ell = 1, \ldots, L, \tag{3}$$

and

$$\underset{\sim}{\gamma}_g = \left[\sum_{\ell=1}^{L} \underset{\sim}{\gamma}^{(\ell)}\right]/L. \tag{4}$$

Since (3) and (4) imply that

$$\sum_{\ell=1}^{L} \underset{\sim}{\Delta}^{(\ell)} = \sum_{\ell=1}^{L} \underset{\sim}{\gamma}^{(\ell)} - \sum_{\ell=1}^{L} \underset{\sim}{\gamma}_g = \underset{\sim}{0}_k,$$

$$\underset{\sim}{\Delta}^{(L)} = -\sum_{\ell=1}^{L-1} \underset{\sim}{\Delta}^{(\ell)}, \tag{5}$$

and we need only consider $\underset{\sim}{\gamma}_g$ and $\underset{\sim}{\gamma}^{(\ell)}$, $\ell = 1, \ldots, L-1$.

The vectors $\underset{\sim}{\gamma}_g$ and $\underset{\sim}{\delta}^{(\ell)}$, $\ell = 1, \ldots, L-1$, may be further decomposed by means of

$$\underset{\sim}{\gamma}_g = \underset{\sim}{X}_0 \underset{\sim}{\nu} \tag{6}$$

and

$$\underset{\sim}{\delta}^{(\ell)} = \underset{\sim}{X}_\ell \underset{\sim}{\mu}^{(\ell)}, \quad \ell = 1, \ldots, L-1, \tag{7}$$

where $\underset{\sim}{X}_0, \ldots, \underset{\sim}{X}_{L-1}$ are $k \times k$ orthonormal matrices and the k-vectors $\underset{\sim}{\nu}$ and $\underset{\sim}{\mu}^{(\ell)}$, $\ell = 1, \ldots, L-1$, are new general and specific parameters respectively. In a log-difference model, m independent linear functions of these parameters may be specified to be zero so long as they are constructed in conformity with the linear constraints of a log-linear model as previously defined.

For a log-linear model with constraints $\underset{\sim}{B}_1 \underset{\sim}{\gamma} = \underset{\sim}{0}_m$, it is necessary that $\underset{\sim}{B}_1 \underset{\sim}{\Delta} = \underset{\sim}{0}_{m \times s}$. *Suppose we wish to specify*

$\underset{\sim}{B}_1^* \begin{bmatrix} \underset{\sim}{\nu} \\ \underset{\sim}{\mu}^{(1)} \\ \vdots \\ \underset{\sim}{\mu}^{(L-1)} \end{bmatrix} = \underset{\sim}{0}_m$. This is equivalent to

$$\underset{\sim}{B}_1^* \underset{\sim}{Q} \underset{\sim}{\gamma} = \underset{\sim}{0}_m, \tag{8}$$

where

$$\underset{\sim}{Q} = \begin{bmatrix} \underset{\sim}{X}_0 & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{X}_1 & \cdots & \underset{\sim}{0} \\ \cdots & \cdots & \cdots & \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{X}_{L-1} \end{bmatrix} \begin{bmatrix} (1/L)\underset{\sim}{I}_k & \cdots & (1/L)\underset{\sim}{I}_k & (1/L)\underset{\sim}{I}_k \\ (L-1)/L\ \underset{\sim}{I}_k & \cdots & (-1/L)\underset{\sim}{I}_k & (-1/L)\underset{\sim}{I}_k \\ \cdots & & \cdots & \cdots \\ (-1/L)\underset{\sim}{I}_k & \cdots & (L-1)/L\underset{\sim}{I}_k & (-1/L)\underset{\sim}{I}_k \end{bmatrix}, \tag{9}$$

so that $\underset{\sim}{B}_1^* \underset{\sim}{Q}$ is equivalent to $\underset{\sim}{B}_1$ and it is necessary to choose $\underset{\sim}{B}_1^*$ such that

$$\underset{\sim}{B}_1^* \underset{\sim}{Q} \underset{\sim}{A}(s) = \underset{\sim}{0}_{m \times s}, \tag{10}$$

In practice one may often take

$$\underset{\sim}{X}_0 = \underset{\sim}{X}_1 = \cdots = \underset{\sim}{X}_{L-1} = \underset{\sim}{X}.$$

Use of orthogonal polynomials in the specification of the matrix $\underset{\sim}{X}$ has proven useful, although we leave the way open for other choices through the use of arbitrary orthonormal matrices, $\underset{\sim}{X}_0, \ldots, \underset{\sim}{X}_{L-1}$. If appropriate orthogonal polynomials are used, elements of $\underset{\sim}{\nu}$ and $\underset{\sim}{\mu}^{(\ell)}$, $\ell = 1, \ldots, L-1$, may be given interpretations, for example, analogous to linear, quadratic, and higher-order trend terms and their interactions in the analysis of variance (Haberman, 1974), and selection of transformed parameters to be taken to be zero for model simplification may proceed as in that situation.

3. Classification Procedures

We formalize the classification problem of Section 1. Following Martin and Bradley (1972), its essential features are:

(i) There are a finite number L of exclusive and exhaustive populations, $\pi^{(1)}, \ldots, \pi^{(L)}$, from which the individual to be classified may arise.

(ii) There are a finite number J of exclusive and exhaustive categories $C_1, \ldots, C_J$ into which individuals are to be classified.

(iii) Samples from each population, or a single sample from all populations, are available.

(iv) An unknown individual $U \in \Pi^{(\ell)}$ with probability $P(\Pi^{(\ell)})$, $\ell = 1, \ldots, L$, is to be classified.

(v) The classification of U is to be made on the basis of its belonging to cell i*.

(vi) The loss entailed by classification of $U \in C_j$, when $U \in \Pi^{(\ell)}$ is $L_j^{(\ell)}$, $j = 1, \ldots, J$, $\ell = 1, \ldots, L$. These losses are taken to be finite, and may be taken to be nonnegative without loss of generality. Conventionally, correct classification is indicated by zero loss.

The minimum risk classification rule is: Classify $U \in C_{j*}$ if $R_{j*}(i^*)$ is a minimum of $\{R_j(i^*), j = 1, \ldots, J\}$, where

$$R_j(i) = \sum_{\ell=1}^{L} L_j^{(\ell)} P(\Pi^{(\ell)}) P(i|\Pi^{(\ell)})/P(i)$$

$$= \sum_{\ell=1}^{L} L_j^{(\ell)} P(i, \Pi^{(\ell)})/P(i)$$

$$= \sum_{\ell=1}^{L} L_j^{(\ell)} P(\Pi^{(\ell)}|i), \quad j = 1, \ldots, J, \; i = 1, \ldots, k, \tag{11}$$

$P(i)$ denotes the probability of state i, $P(i, \Pi^{(\ell)})$ denotes the joint probability of state i and population $\Pi^{(\ell)}$, and $P(i|\Pi^{(\ell)})$ and $P(\Pi^{(\ell)}|i)$ denote conditional probabilities. If the minimum is not unique, one may choose any C_{j*} such that j^* corresponds with any one of the minimizing values of $R_j(i^*)$ and the risk is not affected. The minimum risk is

$$r_{min} = \sum_{i=1}^{k} P(i) \min_j \{R_j(i)\}. \tag{12}$$

The classification rule above may be simplified somewhat through the use of (3). In the single sample case, $R_j(i)$ in (11) may be expressed as

$$R_j(i) = \sum_{\ell=1}^{L} L_j^{(\ell)} \exp \gamma_i^{(\ell)} / P(i)$$

$$= \sum_{\ell=1}^{L} L_j^{(\ell)} \exp \gamma_{gi} \exp \Delta_i^{(\ell)} / P(i)$$

$$= \exp \gamma_{gi} \left\{ \sum_{\ell=1}^{L} L_j^{(\ell)} \exp \Delta_i^{(\ell)} \right\} / P(i). \qquad (13)$$

Clearly the minimum risk classification rule depends solely on the te in brackets in (13), an expression involving log-difference paramete specific to each population only.

Similarly, when independent samples from each population are available, $R_j(i)$ may be expressed as

$$R_j(i) = \sum_{\ell=1}^{L} L_j^{(\ell)} \exp \gamma_i^{(\ell)} P(\Pi^{(\ell)}) / P(i)$$

$$= \sum_{\ell=1}^{L} L_j^{(\ell)} \exp \gamma_{gi} \exp \Delta_i^{(\ell)} P(\Pi^{(\ell)}) / P(i)$$

$$= \exp \gamma_{gi} \left\{ \sum_{\ell=1}^{L} L_j^{(\ell)} \exp \Delta_i^{(\ell)} P(\Pi^{(\ell)}) \right\} / P(i). \qquad (14)$$

This time the minimum risk classification rule depends on the term in brackets in (14), a slightly more complicated expression than (13), in that the probabilities $P(\Pi^{(\ell)})$ are involved. Again the log-difference parameters general to all populations are not involved.

In practice, estimated classification rules are needed. Log-difference modelling will be used and estimators of $\Delta_i^{(\ell)}$ used in place of $\Delta_i^{(\ell)}$ in (13) and (14).

4. Minimum δ and Approximate Minimum δ Estimators

In this section, we find minimum δ and approximate minimum δ estimators following the general procedures of Redman (1981). We also state conditions which are of use for determination of the existence of a minimum δ estimator in sparse data situations. These conditions will be used in the second example of Section 6.

The minimum δ estimator of $\underset{\sim}{\gamma}$ is that point $\tilde{\underset{\sim}{\gamma}}$ in

$$\Gamma(\underset{\sim}{B}_1) = \left\{ \underset{\sim}{\gamma}: \underset{\sim}{B}_1\underset{\sim}{\gamma} = \underset{\sim}{0}_m, \sum_{i=1}^{k} \exp \gamma_i^{(\ell)} = 1, \ell = 1, \ldots, L, s = L, \text{ or } \sum_{\ell=1}^{L}\sum_{i=1}^{k} \exp \gamma_i^{(\ell)} = 1, s = 1 \right\},$$

which minimizes

$$\delta(\underset{\sim}{\gamma}; \underset{\sim}{n}) = \sum_{\ell=1}^{L}\sum_{i=1}^{k} n_i^{(\ell)}(\log p_i^{(\ell)} - \gamma_i^{(\ell)})^2.$$

The function δ is a transformed χ^2-like function, whose use is motivated by the linear nature of some of the constraints on $\underset{\sim}{\gamma}$ expressed in the parameter space $\Gamma(\underset{\sim}{B}_1)$. The minimum δ estimate may be found through solution of the following system of equations:

$$\underset{\sim}{B}_2[\underset{\sim}{N}(\log \underset{\sim}{p} - \underset{\sim}{\gamma}) - \underset{\sim}{y}(\underset{\sim}{\gamma})] = \underset{\sim}{0}_{kL-m}, \quad \underset{\sim}{B}_1\underset{\sim}{\gamma} = \underset{\sim}{0}_m,$$

and (15)

$$\sum_{i=1}^{k} \exp \gamma_i^{(\ell)} = 1, \ell = 1, \ldots, L, \_ = L$$

or

$$\sum_{\ell=1}^{L}\sum_{i=1}^{k} \exp \gamma_i^{(\ell)} = 1, s = 1.$$

Here, $\underset{\sim}{B}_2$ is an orthocomplement of $\underset{\sim}{B}_1$, that is,

$$\begin{bmatrix} \underset{\sim}{B}_1 \\ \underset{\sim}{B}_2 \end{bmatrix} \begin{bmatrix} \underset{\sim}{B}_1 \\ \underset{\sim}{B}_2 \end{bmatrix}' = \underset{\sim}{I}_{kL},$$

$\underset{\sim}{N}$ is the diagonal matrix with entries $n_1^{(1)}, \ldots, n_k^{(1)}, \ldots, n_1^{(L)}, \ldots, n_k^{(L)}$, and $[\underset{\sim}{\gamma}(\underset{\sim}{\gamma})]' = [y_1^{(1)}(\underset{\sim}{\gamma}), \ldots, y_k^{(1)}(\underset{\sim}{\gamma}), \ldots, y_1^{(L)}(\underset{\sim}{\gamma}), \ldots, y_k^{(L)}(\underset{\sim}{\gamma})]$,

$$y_i^{(\ell)}(\underset{\sim}{\gamma}) = \exp \gamma_i^{(\ell)} \sum_{i=1}^{k} n_i^{(\ell)}(\log p_i^{(\ell)} - \gamma_i^{(\ell)}), \quad s = L,$$

and

$$y_i^{(\ell)}(\underset{\sim}{\gamma}) = \exp \gamma_i^{(\ell)} \sum_{\ell=1}^{L} \sum_{i=1}^{k} n_i^{(\ell)}(\log p_i^{(\ell)} - \gamma_i^{(\ell)}), \quad s = 1,$$

$i = 1, \ldots, k, \ell = 1, \ldots, L.$

The approximate minimum δ method is an *ad hoc* variant of the minimum δ method. The approximate minimum δ estimator is denoted by $\tilde{\underset{\sim}{\gamma}}_a$, is easy to compute, and may serve as an initial value for iterative solution of equations (15).

Two steps are required for the calculation of $\tilde{\underset{\sim}{\gamma}}_a$. The first involves minimization of $\delta(\underset{\sim}{\gamma};\underset{\sim}{n})$ over $\{\underset{\sim}{\gamma}: \underset{\sim}{B}_1\underset{\sim}{\gamma} = \underset{\sim}{0}_m\}$. This is the classic minimization of a weighted sum of squares function over an affine space and the set of vectors which yield the desired minimum is given by

$$\bar{\Gamma} = \{\bar{\underset{\sim}{\gamma}}: \bar{\underset{\sim}{\gamma}} = \underset{\sim}{B}_2'[(\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2')^{-}\underset{\sim}{B}_2\underset{\sim}{N} \log \underset{\sim}{p}$$

$$+ (\underset{\sim}{I} - (\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2')^{-}(\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2'))\underset{\sim}{z}], \ \underset{\sim}{z} \in E^{k-m}\}.$$

Note that $(\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2')^{-}$ is a generalized inverse of $\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2'$. When $\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2'$ is nonsingular,

$$\bar{\underset{\sim}{\gamma}} = \underset{\sim}{B}_2'(\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2')^{-1}\underset{\sim}{B}_2\underset{\sim}{N} \log \underset{\sim}{p}. \tag{16}$$

The second step in calculation of $\tilde{\underset{\sim}{\gamma}}_a$ adjusts $\bar{\underset{\sim}{\gamma}}$ so the resultant estimated probabilities sum to one. Thus

$$\tilde{\underset{\sim}{\gamma}}_{ai}^{(\ell)} = \bar{\underset{\sim}{\gamma}}_i^{(\ell)} - c_\ell(\bar{\underset{\sim}{\gamma}}), \quad i = 1, \ldots, k, \quad \ell = 1, \ldots, L, \tag{17}$$

where

$$c_\ell(\bar{\underset{\sim}{\gamma}}) = \sum_{i=1}^{k} \exp \bar{\gamma}_i^{(\ell)}, \quad \ell = 1, \ldots, L, \quad s = L,$$

or

$$c_\ell(\bar{\underset{\sim}{\gamma}}) = c(\bar{\underset{\sim}{\gamma}}) = \sum_{\ell=1}^{L} \sum_{i=1}^{k} \exp \bar{\gamma}_i^{(\ell)}, \quad s = 1.$$

Due to the nature of the set $\bar{\Gamma}$, an approximate minimum δ estimator always exists, and the estimator is unique if and only if $\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2'$ is nonsingular.

The set $\bar{\Gamma}$ is also intimately related to the existence of the minimum δ estimator. Let

$$\bar{\Gamma}^* = \left\{ \bar{\underset{\sim}{\gamma}} : \bar{\underset{\sim}{\gamma}} \epsilon \bar{\Gamma}, \ \sum_{i=1}^{k} \exp \bar{\gamma}_i^{(\ell)} = c_\ell, \ \ell = 1, \ldots, L, \ s = L, \ \sum_{\ell=1}^{L} \sum_{i=1}^{k} \exp \bar{\gamma}_i^{(\ell)} = c, \ s = 1 \right\},$$

where

$$c_\ell = \inf_{\bar{\underset{\sim}{\gamma}} \epsilon \bar{\Gamma}} \{c_\ell(\bar{\underset{\sim}{\gamma}})\} = \inf \left\{ \sum_{i=1}^{k} \exp(\bar{\gamma}_i^{(\ell)}) \right\}, \quad \ell = 1, \ldots, L, \quad s = L,$$

or

$$c = \inf_{\bar{\underset{\sim}{\gamma}} \epsilon \bar{\Gamma}} \{c(\bar{\underset{\sim}{\gamma}})\} = \inf \left\{ \sum_{\ell=1}^{L} \sum_{i=1}^{k} \exp(\bar{\gamma}_i^{(\ell)}) \right\}, \quad s = 1.$$

The set $\bar{\Gamma}^*$ is either empty or a singleton. The following conditions have been obtained by Redman (1981):

Condition 1: A minimum δ estimator exists if and only if $\bar{\Gamma}^*$ is a singleton.

Condition 2: If $(\underset{\sim}{B}_2\underset{\sim}{N}\underset{\sim}{B}_2')$ is nonsingular, then a minimum δ estimator exists.

Condition 3: If all cells contain at least one observation, then a minimum δ estimator exists.

We conclude this section with the remark that, whenever a minimum δ estimator exists, it is unique.

5. Detergent Preference

Our first example involves detergent preference. The data are in Table 1, were collected in a single sample, s = 1, by Ries and Smith (1963), and have been previously analyzed by Goodman (1971), Bishop, Feinberg, and Holland (1975), and others.

The data result from an experiment in which 1008 people expressed their preferences for two brands of detergent, X and M. They responded also to three questions corresponding to three variables:

1. Previous experience with brand M: yes or no,
2. Water hardness: soft, medium, or hard,

and

3. Water temperature: high or low.

Respondents were taken to represent two populations, $\Pi^{(1)}$: consumers who prefer X, and $\Pi^{(2)}$: consumers who prefer M. We assume that the levels of variable 2 are ordered and equally spaced, and we take $\underset{\sim}{X}_0 = \underset{\sim}{X}_1$ with orthonormal columns proportional to the columns of the following matrix:

$$\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 0 & -2 & 1 & 0 & -2 & 1 & 0 & -2 & 0 & -2 \\
1 & 1 & 0 & -2 & -1 & 0 & -2 & -1 & 0 & 2 & 0 & 2 \\
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 \\
1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\
1 & -1 & 0 & -2 & 1 & 0 & 2 & -1 & 0 & -2 & 0 & 2 \\
1 & -1 & 0 & -2 & -1 & 0 & 2 & 1 & 0 & 2 & 0 & -2 \\
1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1
\end{bmatrix}.$$

This matrix has been constructed with the aid of the orthogonal polynomials of Fisher and Yates (1953). Interpretations which may be given to the elements of $\underset{\sim}{\nu}$ and $\underset{\sim}{\mu}^{(1)}$ are given in Table 2.

Table 1. -- Observed and Estimated Frequencies for Ries-Smith Data

	Population $\Pi^{(1)}$								Population $\Pi^{(2)}$				
	Variable*			Observed	Estimated Frequencies				Observed	Estimated Frequencies			
Cell	1	2	3	Frequency	$\hat{\underset{\sim}{Y}}$	$\tilde{\underset{\sim}{Y}}_a$	$\tilde{\underset{\sim}{Y}}$	Cell	Frequency	$\hat{\underset{\sim}{Y}}$	$\tilde{\underset{\sim}{Y}}_a$	$\tilde{\underset{\sim}{Y}}$
1	Y	S	H	19	21.36	21.35	21.35	13	29	28.37	28.43	28.43
2	Y	M	H	57	45.59	45.91	45.91	14	49	60.57	61.14	61.13
3	Y	H	H	23	25.88	26.34	26.34	15	47	34.38	35.09	35.08
4	Y	S	L	47	44.77	44.58	44.59	16	55	59.47	59.38	59.38
5	Y	M	L	24	28.54	28.54	28.54	17	43	37.92	38.01	38.01
6	Y	H	L	37	40.87	40.84	40.84	18	52	54.29	54.39	54.39
7	N	S	H	29	31.06	30.83	30.83	19	27	23.21	23.07	23.07
8	N	M	H	63	66.29	66.29	66.29	20	53	49.55	49.60	49.60
9	N	H	H	33	37.63	38.04	38.04	21	23	28.13	28.47	28.46
10	N	S	L	66	65.10	64.38	64.39	22	50	48.66	48.17	48.18
11	N	M	L	42	41.51	41.21	41.21	23	30	31.03	30.84	30.84
12	N	H	L	68	59.42	58.98	58.98	24	42	44.42	44.13	44.13

*1: Previous Use of Brand M: Y = yes, N = no,
2: Water Hardness: S = soft, M = medium, H = hard,
3: Water Temperature: H = high, L = low.

Table 2. -- Interpretations Which May Be Given to Parameters in Detergent Example

General	Specific	Interpretation
ν_1	$\mu_1^{(1)}$	overall mean
ν_2	$\mu_2^{(1)}$	main effect, var. 1
ν_3	$\mu_3^{(1)}$	linear term, main effect, var. 2
ν_4	$\mu_4^{(1)}$	quadratic term, main effect, var. 2
ν_5	$\mu_5^{(1)}$	main effect, var. 3
ν_6	$\mu_6^{(1)}$	var. 1 by linear term var. 2 interaction
ν_7	$\mu_7^{(1)}$	var. 1 by quadratic term var. 2 interaction
ν_8	$\mu_8^{(1)}$	var. 1 by var. 3 interaction
ν_9	$\mu_9^{(1)}$	linear term var. 2 by var. 3 interaction
ν_{10}	$\mu_{10}^{(1)}$	quadratic term var. 2 by var. 3 interaction
ν_{11}	$\mu_{11}^{(1)}$	var. 1 by linear term var. 2 by var. 3 interaction
ν_{12}	$\mu_{12}^{(1)}$	var. 1 by quadratic term var. 2 by var. 3 interaction

The authors cited have fitted a variety of models to these data. We use the model in which

$$\underset{\sim}{\nu}' = (\nu_1, \ldots, \nu_5, 0, 0, 0, \nu_9, \nu_{10}, 0, 0) = \underset{\sim}{\nu}_0'$$

$$[\underset{\sim}{\mu}^{(1)}]' = [\mu_1^{(1)}, \mu_2^{(1)}, \underset{\sim}{\Omega}_{10}'], \tag{18}$$

because the previous work suggests that this model should fit the data reasonably well and because it involves few parameters, particularly those specific to population $\Pi^{(1)}$.

The estimated frequencies derived from $\hat{\underset{\sim}{\gamma}}$, $\tilde{\underset{\sim}{\gamma}}_a$, and $\tilde{\underset{\sim}{\gamma}}$ are reported in Table 1. Computation of $\hat{\underset{\sim}{\gamma}}$ was effected through use of CONTAB (Zahn, 1974), and computation of $\tilde{\underset{\sim}{\gamma}}_a$ through use of (16) and (17). Computation of $\tilde{\underset{\sim}{\gamma}}$ was effected through solution of equations (15) by means of Newton's Method, (Acton, 1970), with $\tilde{\underset{\sim}{\gamma}}_a$, the initial value. The left-hand side of each of equations (15), evaluated at the first iterate, was less than 10^{-7}, so the first iterate was taken to be the minimum δ estimator.

The goodness-of-fit statistics are $\chi^2(\hat{\underset{\sim}{\gamma}}; \underset{\sim}{n}) = 16.7265$, where $\chi^2(\underset{\sim}{\gamma}; \underset{\sim}{n})$ is the usual Pearson statistic, and $\delta(\tilde{\underset{\sim}{\gamma}}_a; \underset{\sim}{n}) = \delta(\tilde{\underset{\sim}{\gamma}}; \underset{\sim}{n}) = 16.5904$. Under the model, all three statistics are approximately distributed as χ^2_{15} (Redman, 1981). The values of these statistics are only slightly above expectation, indicating good fits of the model by all three methods.

For classification purposes we take populations $\Pi^{(1)}$ and $\Pi^{(2)}$ to coincide with categories C_1 and C_2 and $L_1^{(1)} = L_2^{(2)} = 0$, $L_1^{(2)} = L_2^{(1)} = 1$. For these losses, the classification rule is simplified. Now, we have, from (13),

$$R_1(i^*) \propto \sum_{\ell=1}^{2} L_1^{(\ell)} \exp \Delta_{i^*}^{(\ell)} = \exp \Delta_{i^*}^{(2)} = 1/\exp \Delta_{i^*}^{(1)},$$

and

$$R_2(i^*) \propto \sum_{\ell=1}^{2} L_2^{(\ell)} \exp \Delta_{i^*}^{(\ell)} = \exp \Delta_{i^*}^{(2)},$$

the constants of proportionality being the same. The classification rule reduces to: Classify $U \in C_1$ if $\Delta_{i^*}^{(1)} \geq 0$, and $U \in C_2$ otherwise. For the model used here, we have assumed

$$\underset{\sim}{\Delta}^{(1)} = \underset{\sim}{X}_1 \begin{bmatrix} \mu_1^{(1)} \\ \mu_2^{(1)} \\ \varrho_{10} \end{bmatrix},$$

and the minimum δ estimates of $\mu_1^{(1)}$ and $\mu_2^{(1)}$ are $\tilde{\mu}_1^{(1)} = 0.0043$, $\tilde{\mu}_2^{(1)} = -0.7062$. Thus, the estimated classification rule is: Classify $U \in C_1$ if $\tilde{\Delta}_{i^*}^{(1)} = (0.0043\ X_{1i^*1} - 0.7062\ X_{1i^*2}) \geq 0$, where X_{1ij} denotes the (i, j)-element of $\underset{\sim}{X}_1$, and classify $U \in C_2$ otherwise.

Similarly, the minimum risk in (12) reduces to the minimum probability of misclassification,

$$r_{min} = \sum_{i=1}^{k} P(i) \min\{P(\Pi^{(1)}|i),\ P(\Pi^{(2)}|i)\}.$$

To estimate this probability, we simplify

$$r_{min} = \sum_{i=1}^{k} P(i)\min\{P(i, \Pi^{(1)})/P(i),\ P(i, \Pi^{(2)})/P(i)\}$$

$$= \sum_{i=1}^{k} \min\{P(i, \Pi^{(1)}),\ P(i, \Pi^{(2)})\}, \tag{19}$$

and make use of Table 1, from which estimates of $P(i, \Pi^{(\ell)})$ may be obtained. For instance, for state 1 the minimum δ estimators are

$$\tilde{\pi}_1^{(1)} = \tilde{\pi}_1 = \tilde{P}(1, \Pi^{(1)}) = 21.35/1008 = 0.021,$$

and

$$\tilde{\pi}_1^{(2)} = \tilde{\pi}_{12} = \tilde{P}(1, \Pi^{(2)}) = 28.43/1008 = 0.028.$$

The estimated contribution to (19) for $i = 1$ is 0.021. The estimate of r_{min} is 0.428 for $\tilde{\gamma}$ and $\tilde{\gamma}_a$, it is 0.429 for $\hat{\gamma}$, and, for all three estimators, the proportion of individuals in the study that would be misclassified is 0.429. These figures compare favorably with the estimate of the probability of misclassification 0.421 based on the full model but are by no means impressive.

One may even wonder if inclusion of $\mu_1^{(1)}$ and $\mu_2^{(1)}$ in the model aids in the fit of the data. To examine this, we assume $\gamma = \gamma_0$ and test

$$H_0: [\underset{\sim}{\mu}^{(1)}] = \underset{\sim}{0}_{12},$$

against

$$H_a: [\underset{\sim}{\mu}^{(1)}]' = [\mu_1^{(1)}, \mu_2^{(1)}, \underset{\sim}{0}_{10}'].$$

Under H_0, $\delta(\tilde{\gamma}_1; \underset{\sim}{n}) - \delta(\tilde{\gamma}; \underset{\sim}{n})$ is approximately distributed as χ_2^2, where $\tilde{\gamma}_1$ is the minimum δ estimate of $\tilde{\gamma}$ under H_0. We obtain 20.6404 for this statistic,

a value which indicates rejection of the null hypothesis and suggests that inclusion of $\mu_1^{(1)}$ and $\mu_2^{(1)}$ significantly improves the fit of the model to the data. This test demonstrates that the two populations are indeed different under the restricted model defined in (18). Classification should be possible.

6. Hypoxic Trauma

The second example involves data used by MB[1] and are concerned with history and behavior of infants following hypoxic trauma[2]. For these data $s = 1$, $I = 4$; all variables are dichotomous:

1. race, white or nonwhite,
2. suggestive or nonsuggestive medical history of mother,
3. infant first breath before or after five seconds,
4. infant first cry before or after 30 seconds.

The populations are $\Pi^{(1)}$: Infants with Apgar scores[3] of seven or below and $\Pi^{(2)}$: Infants with normal Apgar scores. The data are in Table 3.

We take $\underset{\sim}{X}_0 = \underset{\sim}{X}_1$ with orthonormal columns proportional to the columns of the following matrix:

$$\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 & 1 \\
1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 \\
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\
1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 \\
1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\
1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1
\end{bmatrix}$$

[1] MB reference unpublished data of Joan C. Martin and Celia Lamoer, Duke University.

[2] Hypoxic trauma: Damage to an infant during or shortly after birth caused by oxygen deficiency.

[3] Apgar score: An index of the level of physiological functioning based on symptoms of the infant observed shortly after birth. See Apgar (1953).

Table 3. -- Observed and Expected Frequencies for Low and Normal Apgar Score Infants

Variable*				Low Apgar Score Group			Normal Apgar Score Group			
					Expected Frequency Based on				Expected Frequency Based on	
1	2	3	4	Cell	Observed Frequency	$\tilde{\gamma}_a$	$\tilde{\gamma}$	Cell	Observed Frequency	$\tilde{\gamma}_a$	$\tilde{\gamma}$
1	1	1	1	1	0	1.40	1.26	17	0	.44	.38
1	1	1	0	2	2	1.75	1.13	18	0	2.74	1.69
1	1	0	1	3	6	4.46	4.77	19	1	1.40	1.36
1	1	0	0	4	15	12.97	14.35	20	23	20.32	21.40
1	0	1	1	5	0	.62	.58	21	0	.36	.31
1	0	1	0	6	0	.82	.52	22	0	2.37	1.47
1	0	0	1	7	2	1.98	2.06	23	1	1.14	1.17
1	0	0	0	8	6	6.11	6.61	24	21	17.54	18.53
0	1	1	1	9	3	2.62	2.50	25	0	.47	.42
0	1	1	0	10	0	4.68	2.91	26	0	4.17	2.56
0	1	0	1	11	9	8.33	8.94	27	0	1.49	1.52
0	1	0	0	12	39	34.68	36.82	28	36	30.88	32.37
0	0	1	1	13	0	1.17	1.09	29	0	.38	.35
0	0	1	0	14	0	2.20	1.34	30	0	3.60	2.22
0	0	0	1	15	4	3.70	3.91	31	2	1.21	1.25
0	0	0	0	16	20	16.33	16.95	32	29	26.66	28.02

*
1: race; 1 = white, 0 = nonwhite
2: 1 = suggestive history of mother
0 = nonsuggestive history of mother
3: 1 = first breath before 5 seconds
0 = first breath after 5 seconds
4: 1 = first cry before 30 seconds
0 = first cry after 30 seconds

This matrix has been derived from MB (Equation 2.1). The parameter associated with the first column of this matrix may be interpreted as an overall mean, those associated with columns 2-5 may be interpreted as main effects of variables 1-4 respectively, and parameters associated with succeeding columns as interactions.

Due to the many empty cells in this data set, the possibility that minimum δ estimators do not exist for many models is of concern. Therefore, a preliminary study to identify a reasonable model for which a minimum δ estimator exists had to be undertaken. The results of this preliminary examination of the data are contained in Table 4. In step 1, it was determined that a minimum δ estimator does not exist for the model in which elements of $\underset{\sim}{\nu}$ and $\underset{\sim}{\mu}^{(1)}$ corresponding to main effects were included (Model 1). If $\mu_4^{(1)}$ is deleted from Model 1, a minimum δ estimator does exist (Model 2). Elements $\mu_1^{(1)}$, $\mu_2^{(1)}$, $\mu_3^{(1)}$, and $\mu_5^{(1)}$ and no other specific terms are included in succeeding models. The existence of a minimum δ estimator for Model 3, which involves elements of $\underset{\sim}{\nu}$ corresponding to main effects and first-order interactions, was checked next. No minimum δ estimator exists for this model, and so, in the final step, the existences of minimum δ estimators for models in which general main effects and three first-order interaction terms involving one of the variables were checked (Models 4-7, variables 1-4).

Minimum δ estimators exist for Models 2, 4, and 7. Approximate minimum δ estimators were calculated for these models and the values of the corresponding $\delta(\tilde{\underset{\sim}{\gamma}}_a; \underset{\sim}{n})$ are given in Table 4. The value of $\delta(\tilde{\underset{\sim}{\gamma}}_a; \underset{\sim}{n})$ for Model 7 appears to be substantially lower than for the other models and so Model 7 was selected as our model.

Table 4. -- Existence of Minimum δ Estimators

Model	General Terms Included	Terms Specific to Pop. 1	Existence	$\delta(\tilde{\gamma}_a; n)$
1	$\nu_1, \ldots, \nu_5$	$\mu_1^{(1)}, \ldots, \mu_5^{(1)}$	No	---
2	$\nu_1, \ldots, \nu_5$	$\mu_1^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}, \mu_5^{(1)}$	Yes	9.27
3	$\nu_1, \ldots, \nu_5, \nu_6, \ldots, \nu_{11}$	$\mu_1^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}, \mu_5^{(1)}$	No	---
4	$\nu_1, \ldots, \nu_5, \nu_6, \nu_7, \nu_8$	$\mu_1^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}, \mu_5^{(1)}$	Yes	11.28
5	$\nu_1, \ldots, \nu_5, \nu_6, \nu_9, \nu_{10}$	$\mu_1^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}, \mu_5^{(1)}$	No	---
6	$\nu_1, \ldots, \nu_5, \nu_7, \nu_9, \nu_{11}$	$\mu_1^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}, \mu_5^{(1)}$	No	---
7	$\nu_1, \ldots, \nu_5, \nu_8, \nu_{10}, \nu_{11}$	$\mu_1^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}, \mu_5^{(1)}$	Yes	5.09

Computation of $\tilde{\gamma}$ was again effected iteratively by means of Newton's Method with use of $\tilde{\gamma}_a$ as an initial estimate. After four iterations, $\delta(\tilde{\gamma}; \underset{\sim}{n}) = 3.2967$ was obtained. Estimated frequencies based on $\tilde{\gamma}$ and $\tilde{\gamma}_a$ are given in Table 3. With the exception of a few zero cells, the observed and expected frequencies appear to agree rather well. Observed and expected frequencies agree in the zero cells more closely for the MB model, but it should be noted that the MB model uses nine more parameters than the present model.

Again we take $L_1^{(1)} = L_2^{(2)} = 0$, $L_1^{(2)} = L_2^{(1)} = 1$. As in the previous example, the classification rule depends only on $\Delta_{i*}^{(1)}$, and the probability of misclassification is given by (19). For Model 7,

$$\underset{\sim}{\Delta}^{(1)} = \underset{\sim}{X}_1 \begin{bmatrix} \mu_1^{(1)} \\ \mu_2^{(1)} \\ \mu_3^{(1)} \\ 0 \\ \mu_5^{(1)} \\ \underset{\sim}{0}_{11} \end{bmatrix},$$

and $\tilde{\mu}_1^{(1)} = 1.0508$, $\tilde{\mu}_2^{(1)} = -0.7479$, $\tilde{\mu}_3^{(1)} = 0.8926$, and $\tilde{\mu}_5^{(1)} = 2.3272$. Thus, the estimated classification rule is: Classify $U \in C_1$ if $\tilde{\Delta}_{i*}^{(1)} = (1.0508\ x_{1i*1} - 0.7479\ x_{1i*2} + 0.8926\ x_{1i*3} + 2.3272\ x_{1i*5}) \geq 0$, and classify $U \in C_2$ otherwise. The estimates of the probability of misclassification are 0.375 for $\tilde{\gamma}_a$, and 0.377 for $\tilde{\gamma}$, and, for both estimators, the proportion of individuals that would be misclassified is 0.379. This proportion matches the proportion misclassified for the MB model.

7. Acknowledgments

Much of this work was prepared as part of the author's Ph.D. dissertation (Redman, 1980) under the direction of Professor Ralph A. Bradley. His patience and dedication are gratefully acknowledged. He is the Principal Investigator of the referenced Office of Naval Research Contracts.

The author wishes also to acknowledge the assistance of Professor Frederick W. Leysieffer who directed the referenced training grant supported by the National Institute of Environmental Health Sciences.

REFERENCES

Acton, F. S. (1970), Numerical Methods that Usually Work, New York: Harper and Row Publishers.

Apgar, V. (1953), "A Proposal for a New Method of Evaluation of the Newborn Infant," Current Researches in Anesthesia and Analgesia, 32, 260-267.

Bishop, Y. M. M., Feinberg, S. E., and Holland, P. W. (1975), Discrete Multivariate Analysis: Theory and Practice, Cambridge: The MIT Press.

Fisher, R. A. and Yates, F. (1963), Statistical Tables for Biological, Agricultural, and Medical Research (6th Edition), New York: Hafner Publishing Company.

Goodman, L. A. (1971), "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classification," Technometrics, 13, 33-61.

Haberman, S. J. (1974), "Log-linear Models for Frequency Tables with Ordered Classifications," Biometrics, 30, 589-600.

Martin, D. C., and Bradley, R. A. (1972), "Probability Models, Estimation, and Classification for Multivariate Dichotomous Populations," Biometrics, 28, 203-222.

Redman, T. C. (1980), Minimum δ Estimation for Log-Linear Models, Ph.D. Dissertation, Florida State University, Strozier Library, Tallahassee, Fla. 32306.

Redman, T. C. (1981), "New Estimation Methods for Log-Linear Models," FSU Statistics Report No. M562, ONR Technical Report No. 151, Department of Statistics, Florida State University, Tallahassee, Fla. 32306.

Ries, P. N. and Smith, H. (1963), "The Use of Chi-Square for Testing in Multidimensional Problems," Chemical Engineering Progress, 59, 39-43.

Zahn, D. A. (1974), "Documentation for CONTAB: A Computer Program to Aid in the Analysis of Multidimensional Contingency Tables Using Log-Linear Models," FSU Technical Report M292, Department of Statistics, Florida State University, Tallahassee, Fla. 32306.

ECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M592 ONR No. 15	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) Use of Log-Linear Models in Classification Problems		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M592
7. AUTHOR(s) Thomas C. Redman		8. CONTRACT OR GRANT NUMBER(s) ONR No. N00014-76-C-0394 ONR No. N00014-80-C-0093
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217		12. REPORT DATE December, 1981 13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

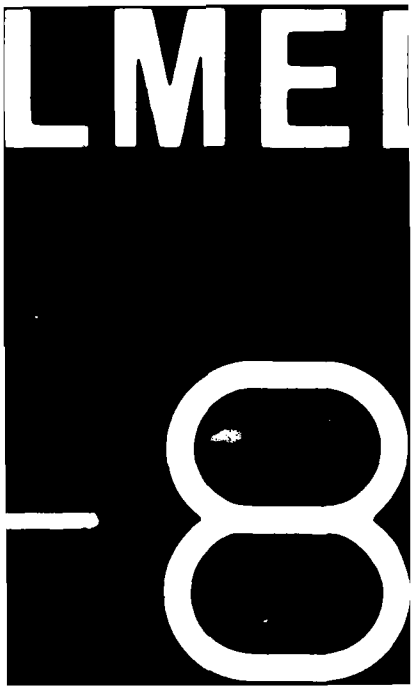
18. SUPPLEMENTARY NOTES

19. KEY WORDS

Classification problem, log-difference models, minimum δ estimation, existence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we consider use of some special log-linear models and minimum δ estimation in the multivariate classification problem posed by Martin and Bradley (1972). We first define these models, called log-difference models, and show that the minimum risk classification rule depends only on a certain subset of the new parameters. We then review minimum δ estimation, in particular the minimum δ estimator, the approximate minimum δ estimator, and their existence properties. Two examples are worked. The first involves detergent preference and illustrates how extensions to the case in which not all variables are dichotomous may be obtained through the use of orthogonal polynomials. The second example involves infant hypoxic trauma, and many cells are empty. The existence conditions are used to find a model for which estimates of cell frequencies exist and are in good agreement with the observed data.


LME
8